CORRECTED
VERSION*

PCT

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(54) Title: SEPARATION, SCREENING, AND IDENTIFICATION OF BIOLOGICAL TARGETS

(57) Abstract

    The present invention relates to the field of proteomics. More specifically, the present invention describes methods and apparatus for the isolation, characterizing, screening, recombining and interacting of biological molecules such as proteins, peptides, nucleic acids and ligands so as to analyze various biological activities of these molecules individually or on a cellular scale. Moreover, the invention relates to the positional mapping of isolated biological molecules in multiple solution–base separation means so as to provide a unique set of identifying characteristics for each biological molecule in a system. The invention further relates to the utilization of this information for the simultaneous screening, selection and enrichment of interactive ligands, substrates or other interactive molecules in many thousands of parallel ligand–target, substrate–enzyme or other biological interactions. The invention further relates to identification and display of the target molecules or interactive molecules for subsequent analysis. The present invention is valuable in the screening and study of potential small therapeutic molecules and their interactions in various cell types of choice.

SEPARATION, SCREENING, AND IDENTIFICATION
OF BIOLOGICAL TARGETS

5   **TECHNICAL FIELD AND INDUSTRIAL APPLICATION OF INVENTION**

The invention relates to a method and apparatus for the isolation, characterizing, screening, recombining and interacting of biological molecules such as proteins, peptides, nucleic acids and ligands so as to analyze various biological

10   activities of these molecules individually or on a global scale. Moreover the invention relates to the positional mapping of isolated biological molecules in multiple solution-based separation means so as to provide a unique set of identifying characteristics for each biological molecule. The invention further relates to the utilization of this information for the simultaneous screening, selection and enrichment of interactive

15   ligands, substrates or other interactive molecules in many thousands of parallel ligand-target, substrate-enzyme or other biological interactions. The invention further relates to identification and display of the target molecules or interactive molecules for subsequent analysis.

20   **BACKGROUND OF THE INVENTION**

The science of molecular biology has attempted to understand how the complex systems found in living organisms function by borrowing a successful strategy from other physical science fields. Physicists, for example, have synthesized

25   · global theories through reductionism wherein the individual elements of a system were first understood separately and then recombined and synthesized into an all-encompassing theory that holds true for a broad range of phenomena and scales. Molecular biology has been quite successful at the first part of this strategy and understands how many individual elements of a living organism function. The process

30   of recombining this vast amount of information in order to synthesize an overarching or global understanding of how living organisms function has so far been limited.

The primary reason for this situation is that even the simplest biological systems, are extremely complex adaptive systems that exhibit emergent

phenomenological behavior on the global scale that cannot be predicted from understanding the individual elements such as enzymes and signaling molecules. Current attempts to gain a more global understanding of living organisms is focused in the Human Genome Project, but it is understood that the mere listing of all the

5   genes in the human genome reveals little about how they interact, anymore than a parts list for a 747 jet-liner would tell us how to assemble one without an assembly diagram. A related science called Functional Genomics, (US patent No. 5,695,937 Kinzler, K et al. 1997) for instance, attempts to measure when, how and where these genes are expressed in living cells.

10

Functional Genomics information, however, only provides a general idea of the biological state of a living cell because the protein products of gene expression can be altered in many ways after they have been expressed through interactions with other protein products. It is a well known but often ignored fact, that the best model for

15 understanding the complex functioning of a living cell and observing its biological activity state at some moment in time, is at the whole systems level where all of the protein products and their genes interact.

There are very few global scale methods for analyzing native proteins. Two

20 dimensional gel electrophoresis is one of these global techniques for analyzing multiple proteins. Recent research, using two dimensional gel electrophoresis, has attempted to analyze the predictability of the gene expression data from Functional Genomics for determining protein mass in a whole cell. A good discussion of how comprehensive these current 2D gel methods are at measuring all of the expressed

25 proteins in a cell at a given moment is in "Correlation between Protein and mRNA Abundance in Yeast" Gygi et al. Molecular and Cellular Biology 19:3 p1720-1730 (1999). The article demonstrates that in the eukaryotic organism S. cerevisiae (bakers yeast), for which all 6000 genes have been determined, gene expression data measured as mRNA copies, shows a reasonable correlation to protein mass for only

30 the 10-20 most abundant protein products. All other proteins expressed in the yeast organism vary in quantity from the gene expression level by orders of magnitude. The biological activity state of these proteins is even less predicted by mRNA levels. Furthermore, no more than 10-20% of these 6000 genes are detectable by the present application of 2D-gel electrophoresis. It is expected that this situation remains

35 true for all other eukaryotic cells.

Because of the problem discussed above, it is a primary goal of the science of molecular biology and the drug discovery industry to achieve a comprehensive and global analysis of all of the proteins in a living cell. New techniques, especially those involving mass spectroscopic analysis, are providing fast and sensitive methods for pairwise comparison of the relative abundance of a given protein under different conditions (Nature Biotechnology *17*, 994-999 (1999) for example), but do not give a global picture of biological activity changes.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide several means of comprehensive and global analysis of not only the abundance, but also the in vitro or ex-vivo (freshly broken cell) biological activity state of proteins from a living cell at some moment in time. The means of analysis, in the current invention, depends on a predetermined positional mapping in multiple separation parameters down to the lowest abundance, of all the proteins in a cell. These multiple separation parameters, including the various forms of chromatographic separation, do not resolve (isolate) the proteins explicitly, but provide for implicit resolution based on a unique combination of positional data for each protein from all the multiple separation parameters taken as a whole. The present invention provides such a means of analysis.

It is also an object of the invention to provide a means for the exact and comprehensive positional mapping of every protein in a naturally occurring aggregate such as a cell lysate. This said means allows for the selective recombining by pooling of semi-purified cellular fractions profiled on the above described multiple separation means. These recombined pools are formulated using the positional mapping information so as to include or completely exclude certain specific target proteins from the recombined pool.

An additional object of the present invention is to provide a means, using said exact positional mapping, for the systematic deconvolution of the individual sources of biological activity measured in these same semi-purified cell protein fractions. This systematic deconvolution of individual sources of biological activity is particularly

useful when the biological activity is labile and must be measured immediately after lysing a cell.

5   The ability to implicitly isolate the biological activity of individual target molecules or mixtures of molecules by pooling semi-purified cell protein fractions so as to partially enhance (include) or completely eliminate a given target (exclude) is another key technology of the present invention. This key technology is dependent on complete positional mapping in multiple separation parameters, down to the very lowest abundance, of all proteins in a cellular inventory, which is provided for in the

10   present invention.

In the present invention the nature of the targeted biological activity can be any natural interaction or enzymatic activity of the target molecule, or it can be an artificially created interaction not generally found in a living cell, such as

15   pharmaceutical activity, immunological or artificial binding epitope interaction, or selective artificial enzymatic substrates. A further object of the present invention is to provide a means of scoring, selecting, screening and characterizing said biological activity interactions with many thousands of the target molecules in a cell in parallel and simultaneously so that the biological activities of an array of many target molecule

20   interactions can be analyzed without further physical separation and without interference with one another.

For example, a large set of selectively recombined cellular fractions, created in pairs for each target molecule in the target set, can be analyzed using any solution

25   based biological assay. Since the biologically active molecules in question would be widely distributed in the various reformulated pools, a modestly positive signal in both the partially enhanced and eliminated pools of some tested pair would not be significant. However, if a pair of said selectively recombined cellular fractions were tested and showed a pattern of increased activity, above that in other pairs in the

30   enhanced pool, and was completely devoid of activity in the pool that completely eliminated a given target, it would indicate a strong correlation between the measured biological activity and the specific target molecule associated with the pair of recombined cellular pools so tested.

35   The present invention provides a means of analysis for natural interactions,

among them are biologically significant protein-protein interactions or protein (enzyme)-substrate interactions involved in normal biological activities in living cells. This said method of analysis of the present invention provides for interactions in a native state or conformation, in contrast to the existing art known as two-hybrid or

5   three-hybrid methods (US patent No. 5,283,173 Fields, S. et al. 1994, US patent No. 5,928,868 Liu, Jun et al. 1999) wherein some proximity signal between interacting species is detected. The existing two and three hybrid methods generally involve the artificial over-expression of the proteins to be tested in recombinant expression vectors, such as yeast cells genetically engineered to contain the test protein. These

10 existing two and three hybrid methods represent, at best, a surrogate model of protein-protein interaction far removed from the actual conditions of interaction between native proteins in a living cell.

In a related art of artifactual biological interactions, such as epitope binding or

15 enzymatic activity towards a synthetic substrate, several recent methods (US patent 5,837,500 Ladner et al. et al 1998.; US patent 5,338,665 Schatz et al. 1994, US patent No. 5,565,332 Hoogenboom et al. 1996) and most notably (US patent No. 5,723.323 and US patent No. 5,824,514, Kauffman et al. 1998) provide for the selection and directed molecular evolution of interacting molecules from a very large

20 stochastically-generated collection of candidate molecules, whose identity and structure is somehow encoded in the molecule or is traceable to the molecule. Said collections of stochastically-generated candidate molecules are generally referred to as "libraries of molecules" or "stochastic libraries". Examples include a wide variety of surface display expression vectors systems that allow for the clonal expansion of

25 candidate vectors that interact with the target protein (see, for instance; US patent 5,338,665 Schatz, P et al.1994 and Yeast surface display for screening combinatorial polypeptide libraries, Border, E et al. *Nature Biotechnology* 15, p 553 (1997).

It is a further object of the present invention to provide a means to extend

30 these methods through parallel selection of many thousands of candidate interactive molecules with many or all the target proteins in a cellular inventory simultaneously. The parallel nature of this directed molecular evolution, which we will hereto refer to as "directed molecular co-evolution", allows candidate molecules to be chosen and selected for each interactive target molecule on the basis that they are exceptionally

35 specific for the target molecule and comparatively non-interactive to any other target

molecule in the cellular inventory.

Following many stages of parallel selection, which we will hereto refer to as "generations", through directed molecular co-evolution of stochastically generated
5   candidate molecules, a highly evolved compendium of artificial interactive molecules is created that can interact together as a complex adapted system with the entire target molecule inventory without further segregation or purification. In this parallel selection process, specificity in the presence of all other target molecules is a more important criterion of selection than the strength of the interaction. Several means of
10   scoring said stochastically generated candidate molecules for particular criteria are provided in the present invention.

Such a compendium of co-evolved molecules can subsequently be positionally arrayed either in solution compartments or immobilized to a surface so as to analyze
15   changes in target protein interactions on a global scale. Said compendium of co-evolved molecules may be isolated molecules or may be displayed on the surface of the said surface display expression vectors that created them. This compendium of co-evolved molecules can also be used to create exceptionally specific intracellular tags for the corresponding target molecules (within living cells) following chemical
20   conjugation to a desired functional chemical moiety.

In the prior art of directed molecular evolution, the characteristic typically selected for is binding to a target (US Patent 5,403,484 Ladner et al 1993 for example). It is a further object of the present invention to provide a means of scoring
25   the interaction for selection utilizing the method of selective recombination by pooling of cellular fractions described above as a test target and causing a non-denatured potentially active form of the target to be present in enhanced abundance or completely eliminated from a test cellular pool. This object of the present invention increases the nature of the selectable characteristics used in directed molecular
30   evolution of interacting molecules to include any measurable biological activity of the target molecule, not just binding, as is common in the existing art.

This unique parallel selection of many different interactive molecules for high specificity in the presence of an entire cellular inventory of proteins first requires a
35   means for selecting many different initial candidate molecules for each member of the

target set in a systematic manner. The direct competition of many unrelated candidate molecules in an initial stochastic library would provide a better chance of selecting candidate molecules that would function together. By analogy to natural selection, having many "initial leads" will result in many more "evolutionary lines of competitors"

5   for each said target interaction. The standard method of selecting interactive molecules through a process of stepwise enrichment based on affinity to a single immobilized target, often referred to as bio-panning (US patent No. 5,403,484 Ladner 1993), is not suitable for this purpose. A related technique (US patent No. 5,514,548 Krebber, K et al. 1996) involving enhanced infectivity and expansion of a bound

10   display vector is also of limited value for this application.

In contrast, a novel means of presenting the target set, or molecular fragments from the target set, to all the stochastically generated candidate molecules is another object provided in the current invention. In this initial selection method a series of

15   target molecules are immobilized on particles and transferred in a serial fashion from one subset of the stochastically generated molecules to another in such a fashion that relatively weak interactions result in isolating the interacting target/candidate pair.

Once identified, subsequent stages of molecular selection require the scoring

20   of many competing interactive candidate molecules for subtle differences in their interaction with the non-target molecules in the target set. The present invention includes another novel method of selecting and ranking closely related specifically interacting molecules for interaction with a large pool of target or non-target molecules, in this case having only subtle differences in said target or non-target

25   interactions.

The present invention relates to a series of methods and apparatus that first determine the positional coordinates of a very large number of naturally occurring or naturally grouped biological molecules such as the cellular inventory of expressed

30   proteins in a living cell at one moment in time, hereafter referred to as the target set. These said positional coordinates are identified for each biological molecule in the target set on a plurality of solution based separation means that distribute the said biological molecules in complimentary (orthogonal) patterns such that a given biological molecule's positional coordinates are unique.

35

The present invention further relates to a method of using this positional coordinate information to recombine a plurality of isolated fractions from a plurality of said solution based separation means in such a fashion that two carefully formulated pools are created for each target molecule in the target set.

5

In one embodiment, those said fractions from each solution based separation means that contain a given target molecule are recombined to create a subset pool of the entire target set containing the given target in an enriched manner. Many other target molecules that co-elute in the same said fractions from the same said solution

10 based separation means will be incidentally included.

In a second embodiment, those said fractions from each solution based separation means that do not contain any trace of the given target molecule are recombined to create a subset pool of the entire target set that incidentally includes

15 every molecule in the target set except the target molecule in question. Because of the complimentary (orthogonal) distribution of target molecules between the various solution based separation means, those members of the target set incidentally removed along with the given target molecule from one such solution based separation means will be incidentally included from one or more of the other solution

20 based separation means.

This process is repeated for each said target molecule in the target set. The recombining process described herein is informationally complex but physically simple and can be accomplished using existing robotic fluid handling apparatuses. The pairs

25 of recombined subsets of pooled fractions of the entire target set as in the first described case wherein the given target molecule is enhanced, will hereinafter be called the "inclusionary pool". The pairs in the second described case wherein the given target molecule is excluded, will hereinafter be called the "exclusionary pool". The pair of inclusionary and exclusionary pools, which contain stable biologically

30 active or non-denatured molecules in solution, provide a powerful general analytical tool for correlating measured biological activity to the target molecule in question and they have many applications within the current invention. It must be stressed that comprehensive positional information for every analyte in the target set is required, because any low abundance undiscovered target molecules would create ambiguity in

35 the individual inclusionary and exclusionary pools.

A further means is provided, using said positional coordinate information from solution based separation means, for the correlation of labile biological activity that can only be measured in whole cell lysates immediately after the breaking or lysing of

5   the cells. In this method the labile biological activity is rapidly profiled and calibrated to fractions of the cell lysate separated in parallel on each of the same said solution based separation means used to create the positional coordinate information. The distribution of the biological activity on each profile is then mathematically fitted to the distribution of each target molecule in order to provide a weighted measure of

10   correlation between the said biological activity and each said target molecule or molecules in the target set.

A correlated subset of potential target molecules is created for each separation means. By mathematically calculating the weighted Boolean intercept of

15   the various said correlated subsets, one or a few target molecules consistently associated with said labile biological activity will emerge as candidates for the same said biological activity.

The aforementioned inclusionary and exclusionary pairs provide a further

20   means of correlation of any stable biological activity that can be measured in vitro in the same said inclusionary and exclusionary pairs. This activity can include any biological activity measured against an exogenous substrate, pharmaceutical compound or biological molecule.

25   In addition to the measurement of biological activity against an exogenous reagent, the method provides for the pair-wise screening of endogenous target molecules for biologically significant target-target interactions by further combinatorial recombination of fractions from the aforementioned solution based separation means. In this embodiment, pooling fractions so as to screen each target molecule for

30   biological interaction against all the other target molecules would require the creation of an exceptionally large number of doubly inclusionary and exclusionary pairs. In this case every possible pair of target molecules would be recombined two at a time so that both targets were enhanced and excluded respectively. If the total number of target molecules were 10^3 then there would be 499,500 such double pairs. Using

35   several rounds of stringency, starting with very broad fractional inclusionary and

exclusionary subsets that do not specifically include or exclude only a single pair of target molecules, one can narrow the number of pair wise interactions that need to be screened, moving to narrower subsets that do contain a single doubly inclusionary and exclusionary pair only for those lower stringency subsets that provide positive

5   interactions.

A means for selecting and screening stochastically generated candidate molecules from stochastic libraries that interact with said exclusionary and inclusionary pools is provided. A stochastic library of candidate molecules expressed

10   on the surface of an expression vector (such as a phage, bacterial cell or yeast cell for instance) is first pre-screened for interaction, to a predetermined level of stringency, against the exclusionary pool (i.e. formulated so as to exclude a particular target molecule) by some means in which the interacting candidate molecule expression vectors are retained or otherwise identified. This pre-screening provides a method by

15   which candidate molecules that are potentially highly specific for the target molecule are enriched in the sub-set of the stochastic library that is not retained. Said enriched sub-set of the stochastic library is subsequently screened for interaction against the inclusionary pool. Retained or otherwise identified candidate molecule vectors are potential candidate molecules for said interaction. This process of exclusionary pre-

20   screening and subsequent inclusionary screening can be repeated for every exclusionary/ inclusionary pair within a target set. The stochastic library can be interacted with said target set in parallel or in a serial fashion.

A further means for selecting and screening closely related stochastically

25   generated candidate molecules from stochastic libraries that interact with said exclusionary and inclusionary pools is provided. The process of directed molecular evolution involves biological descent (in this case the clonal expansion of candidate molecule vectors) with variation. This results in many closely related but variant candidate molecules expressed on the surface of their vectors. Each individual vector

30   has only the candidate molecule corresponding to its variant recombinant gene.

A means of scoring the relative interaction of said related candidate molecules against a chosen target and additionally, scoring the relative interaction of said related candidate molecules for unwanted interaction with all other target molecules in a

35   target set is a desirable goal in directed molecular evolution. A means of said scoring

of relative interaction among closely related variant candidate molecules is provided by attachment of said related candidate molecules or the vectors that display said related candidate molecules to a magnetic particle in the nanometer to micrometer range of sizes.

5

Said magnetic particles are allowed to interact in free solution with the aforementioned exclusionary or inclusionary pools so as to bind those target molecules that interact with said related candidate molecules. This includes very weak interactions with target molecules in the exclusionary pool. Said magnetic particles

10 along with any attached target molecules, are subsequently placed in a mobile solution phase and drawn through a stationary phase medium by application of a magnetic field, thus providing a method for the partitioning of the closely related candidate molecules attached directly or indirectly to the magnetic particles.

15 Said method for the partitioning of related candidate molecules may involve steric hindrance of the nanometer sized magnetic particles through a microporous matrix due to the presence of attached interacting target molecules. A further means of partitioning related candidate molecules is the competitive interaction of target molecules in solution with magnetic particles having related candidate molecules

20 attached and a stationary matrix having relatively weakly interacting candidate molecules or their vectors immobilized on the matrix surface. For example, closely related variant candidate molecules that do not interact with the exclusionary pool (i.e. are highly selective for a given target) can be scored and identified by the rate of magnetic movement through said stationary matrix. The subset of closely related

25 candidate molecules that travel through the stationary matrix with the least interaction will elute first and represent a subset of closely related variants selected for the trait of non-interaction with the exclusionary pool of target molecules.

In a preferred embodiment, a final means is provided for the identification of

30 initial candidate molecules from a stochastically generated library of surface display expression vectors wherein the target molecules or fragments of the target molecules in a purified form are covalently attached to nanometer-scale magnetic particles having a smaller size distinguishable from the size of the vectors in said library. The complete stochastic library of candidate molecules is subdivided during its creation

35 into a plurality of sub-libraries, a plurality of compartments or wells. Each sub-library

contains a very large number of independently stochastically generated candidate molecules. Due to the astronomical number of potential candidate molecules generated in any stochastic process, there will be few, if any, identical candidate molecules in different sub-libraries. The size and number of candidate molecules in

5   each said sub-library is chosen in order to limit the number of potential initial interactions with a target molecule or set of mixed target molecules. It is desired that less than one positive interaction be recorded during the incubation of each said sub-library with one or more magnetically immobilized target molecules.

10   Said magnetically immobilized target molecules are magnetically drawn into a sub-library and incubated to achieve equilibration of any potential interaction. A microporous screen or sieve is placed over the surface of the compartment containing the sub-library. Said screen or sieve provides a means of retaining all candidate vectors, but allowing the passage of said smaller magnetic particles. Magnetic

15   particles interacting with some candidate molecule on the surface of an expression vector would be retained, while all other magnetic particles were magnetically drawn through the screen or sieve into a new sub-library. In one embodiment of the current invention, retained magnetic particles can be detected in the sub-libraries by sensitive magnetic detectors such as super-conducting quantum interference devices. The set

20   of magnetically immobilized target molecules can be serially passed in this fashion from one sub-library to another.

BRIEF DESCRIPTION OF THE FIGURES

25   Comprehension of the invention is facilitated by reading the following description in conjunction with the annexed figures in which:

FIG. 1A is a schematic representation depicting a typical solution based separation means time based elution showing a typical profile (for descriptive

30   purposes only) and the position of a plurality of fractions chosen to correspond and calibrate to fractions in a predetermined reference database which is also schematically depicted showing positional information of typical target molecules.

FIG. 1B is a schematic representation depicting a method of combining

35   fractions from a plurality of time based elution profiles from solution based separation

means that exclude a particular target molecule based on positional information in a predetermined reference database.

FIG. 1C   is a schematic representation depicting a method of combining fractions from a plurality of time based elution profiles from solution based separation means that include a particular target molecule based on positional information in a predetermined reference database.

FIG. 2 is a schematic representation depicting an apparatus and method for the parallel introduction and separation of a sample on a plurality of solution based separation means and a plurality of fractions in arrays, for collecting the time based elution from said separation means, that correspond to the position of fractions in said predetermined reference database.

FIG. 3A is a schematic representation depicting a plotting of one typical biological activity profile as measured in said plurality of fractions in an array and a schematic representation of corresponding fractions in a pre-determined reference database showing corresponding target molecules and a vertical measure of their fit to the biological activity profile. Hypothetical target molecules are labeled A through H, while unlabeled target molecules represent non-correlated target molecules.

FIG. 3B is a schematic representation depicting a plotting of a second different profile of the same biological activity as measured in said plurality of fractions in a second array and a schematic representation of corresponding fractions in a pre-determined reference database showing corresponding target molecules and a vertical measure of their fit to the biological activity profile. Hypothetical target molecules are labeled A, C, D, E, G, H, K, and J, while unlabeled target molecules represent non-correlated target molecules.

FIG. 3C is a schematic representation depicting a plotting of a third different profile of the same biological activity as measured in said plurality of fractions in a third array and a schematic representation of corresponding fractions in a pre-determined reference database showing corresponding target molecules and a vertical measure of their fit to the biological activity profile. Hypothetical target molecules are labeled A, E, F, H, J, K, L, and M while unlabeled target molecules

represent non-correlated target molecules.

FIG. **3D** is a schematic representation depicting a plotting of the intersection of the subset of target molecules in the first subset with the subsets of target molecules
5 in the other subsets, resulting in a best fit target. Hypothetical target molecules are labeled **A, B, C, D, E, F, G, H, J, K, L** and **M** while unlabeled target molecules ·represent non-correlated target molecules, with **H** representing the best fit target.

10 FIG. **4A** is a schematic representation partially depicting a micro-array plate containing sub-libraries of a plurality of stochastically generated surface display expression vectors showing an incubation with target molecules immobilized on paramagnetic particles.

15 FIG. **4B** is a schematic representation partially depicting a micro-array plate containing sub-libraries, a microporous screen and a second micro-array plate containing additional sub-libraries.

FIG. **4C** is a schematic representation depicting a magnetic force field applied
20 perpendicularly to the assembly depicted in FIG. **4B** showing the movement and retention of certain paramagnetic particles.

FIG. **4D** is a schematic representation partially depicting the first micro-array plate containing paramagnetic particles bound to a stochastically generated surface
25 display expression vector, and a removed assembly of a microporous screen and a second micro-array plate containing unbound paramagnetic particles.

FIG. **4E** is a schematic representation partially depicting the first micro-array plate with non-interacting surface display expression vectors, a magnetic force field
30 applied perpendicularly to the assembly and an aligned empty third micro-array plate showing magnetically transferred paramagnetic particles bound to stochastically generated surface display expression vectors.

FIG. **4F** is a schematic representation partially depicting the third micro-array
35 plate with the magnetically transferred paramagnetic particles bound to stochastically

generated surface display expression vectors showing a schematic representation of a magnetometer for detecting paramagnetic particles.

DETAILED DESCRIPTION AND PREFERRED EMBODIMENTS

5

POSITIONAL MAPPING OF TARGET MOLECULES FROM A CELLULAR LYSATE TARGET
SET

In one embodiment, positional coordinate information is determined by the
10   multi-dimensional analysis of a plurality of adjacent fractions along each said solution based separation means of a plurality of such solution based separation means using a large quantity of cell lysate as a reference target set sample. A single cell type or sub-cellular organelle is used for each positional coordinate mapping target set.

15   The outflowing analyte stream from each solution based separation means is then divided into many individual fractions according to elution time. Each individual fraction provides a subset of the full target set that contains many individual biological molecules. Said fractions will contain a substantial number of biological molecules that are common to the adjacent fractions in the elution profile providing a contiguous
20   pattern over the total profile.

Considering the low abundance of some particular biological molecules in target sets (such as the complete protein inventory of a cell), a large mass sample of the target set molecules must first be obtained so that the detection limits of the positional
25   mapping method do not miss the lowest abundance biological molecules. In a preferred embodiment, the strategy for this positional mapping involves the pre-fractionation by some sub-cellular isolation means (i.e. nucleus, cytosol, mitochondria etc.), or by some physiochemical criteria such as molecular weight range or isoelectric point range. This allows a sufficient amount of mass to be loaded on said solution-based separation
30   means to detect the lowest abundance biological molecules in the target set.

These fractions are subsequently analyzed by a method that is able to separate all biological molecules in the fraction to near baseline. Said analysis method typically would be a multi-dimensional separation method such as 2D gel

electrophoresis or one of the many new "hyphenated" methods of on line analysis such as LC-MS-MS (liquid chromatography-tandem quadrapole mass spectroscopy) as described in "Identifying the major proteome components of *Haemophilus influenzae* type strain NCTC 8143" Link, A et al. Electrophoresis *18* p1314-1334

5   (1997). Another method is CIEF-ESI-ICR-MS (capillary isoelectric focusing–electrospray ionization-ion cyclotron resonance mass spectroscopy) described in; "Probing proteomes using capillary isoelectric focusing-electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry" Jensen, PK et al. Analytical Chemistry *71(11)* p2076-2084 (1999).

10

2D gel electrophoresis was first introduced by P.H. O'Farrel in "High-Resolution Two-dimensional Electrophoresis of Proteins" Journal of Biological Chemistry *250*, 4007-4021(1975) and extensively described in "Two Dimensional Electrophoresis" L. Anderson, Large Scale Biology Press, Rockville, MD (1991). A

15   further means of automation of the process of production of suitable 2D acrylamide gels for said analysis is presented in, "Continuous Gel Casting Method and Apparatus" Champage, J. USPTO application no. 09/136,525 filed 19 August 1998, and said application is hereby incorporated by reference.

20   A separate multi-dimensional analysis is thus provided for every fraction of a plurality of adjacent fractions on the elution profiles of a plurality of said solution based separation means. Because said adjacent fractions contain many of the same overlapping target molecules, a contiguous mapping of the target molecules along the profile can be calculated by combining the information from each said fraction

25   analysis using prior art imaging techniques such as automated serial segmentation of stacked images.

The exact position of all said protein analytes are thus determined along the said plurality of separation means elution profiles in the elution dimension of the

30   separating means. The combined result is a detailed mapping of the positional distribution of every protein analyte in said cell lysate reference sample in multiple dimensions of solution based separation. Said plurality of solution based separation means are chosen to include separation means that are highly complimentary to one another i.e. orthogonal in their separation parameters and thus provide positional

35   distributions of protein analytes that are considerably different from one another. Each

protein analyte thus is identifiable by a plurality of positional coordinates in multiple dimensions of solution based separation.

The aforementioned solution based separation means include but are not 5 limited to strong and weak cation exchange chromatography, strong and weak anion exchange chromatography, size exclusion chromatography, hydrophobic interaction chromatography, hydrophilic interaction chromatography, hydroxyapatite chromatography, capillary gel electrophoresis, dye interaction chromatography, fast performance liquid chromatography, reverse phase chromatography, perfusion 10 chromatography and low stringency or non-specific affinity chromatography. These methods are well described in the prior art.

An alternate embodiment of the present invention includes a means of multi-dimensional analysis of a plurality of adjacent fractions on a plurality of solution based 15 separation means and is provided by a preferred method of said analysis described in the concurrently filed provisional patent application titled, "A Multi-channel Method and apparatus for solution based Separation and Detection of Amphoteric Substances in Two Dimensions". Said Application (0126-0009) was filed 16 November 1999, and is hereby incorporated by reference.
20

In said continuously flowing multi-dimensional analysis alternative, each said adjacent fraction on each said solution based separation means represents a single sample for multidimensional analysis and the positional information of individual protein analytes is correlated between adjacent fractions as described above for 2D gel 25 electrophoresis analysis.

EXAMPLE 1

CONSTRUCTION OF RECOMBINED INCLUSIONARY/ EXCLUSIONARY
30   POOLS

Once the aforementioned reference database is determined a sample with a complete cellular inventory or some sub-fraction containing the referenced target set is applied to each said solution based separation means under the same conditions

as were used to create said reference database and separated so as to be calibrated to the reference data base with regard to the position of each target molecule. A plurality of fractions of said profiles of said solution based separation means is collected in a manner that protects and preserves biological activity. Said methods of

5   collection include but are not limited to chilled fraction collection in a biological activity stabilizing buffer and or lyophilization with storage below 0° C. In FIG.1A, the position and order of said plurality of fractions is depicted as calibrated to the position and order of fractions used to create the reference database in such a manner that the inclusion or exclusion of target molecules can be determined for every fraction.

10

While maintaining conditions that protect and preserve biological activity, an automated liquid handling apparatus is used to withdraw a portion of particular fractions and to transfer and deposit said portion of particular fractions into a new container or tube. Said automated liquid handling apparatus is programmed using the

15   information provided by said reference database to transfer and create as shown in FIG. 1B a pool (1) in the first case, containing only those fractions (2) from all said profiles that do not contain a particular target molecule and in the second case, FIG. 1C a pool (3) containing only those fractions (4) that do contain said particular target molecule. The resultant pools (1,3) of a portion of each said fraction from each said

20   profile are referred to as the exclusionary pool (1) and the inclusionary pool (3) to the particular target molecule respectively.

Said automated liquid handling apparatus is programmed to repeat the above described procedure of collection, transfer and pooling for every target molecule in

25   said defined target set. The fractional amount taken of each fraction is inversely proportional on the number of target molecules in the target set. The resultant exclusionary and inclusionary pools are either immediately sub-divided into many smaller pools in an array of wells or spotted onto a solid phase or the pools are stored for subsequent sub-division under said conditions that protect and preserve biological

30   activity.

The final result is a complex array of biologically active target molecules that can be assayed as a whole system for any particular biological activity that can be measured in solution. In general, a particular biological target molecule responsible in

35   whole or part for the particular assayed biological activity will be incidentally located,

along with many other target molecules, in most of the wells or spotted positions of said array. Moderate and comparable activity in both pools of an exclusionary and inclusionary pair does not indicate correlation of the said biological activity to that corresponding target molecule.

Taken as a whole, these comparable signal pairs define a baseline measurement of biological activity that can be used to determine when one or more pairs of exclusionary and inclusionary pools demonstrate a complete lack of biological activity and an enhanced biological activity. Analysis of the position of said enhanced and eliminated biological activity in the array would indicate the identity within the reference database of the responsible target molecule. Subsequent correlation of the responsible target molecule in the reference database to known target molecules in existing molecular databases would provide absolute identification of the target responsible for the measured biological activity.

EXAMPLE 2

CORRELATION OF LABILE BIOLOGICAL ACTIVITY IN CELLULAR LYSATES TO TARGET MOLECULES

There are many measurable biological activity assays that are stable with time or can preserve said biological activity for a period of time under suitable conditions. For example the biological activity of an allosteric isoenzyme such as alcohol dehydrogenase 2 (ADH2) in the yeast S. cerevisiae, which catalyses the conversion of ethanol into acetaldehyde, is stable in solution for many hours. There are many other biologically measurable activities that can only be measured in living cells or in solution for a short time after the lysing of the cell. Many such labile activities involve biological activities that require physical interaction by several loosely held subunits that can diffuse away from one another after lysing the cell into a buffering solution. Examples include the kinase activity of many enzymes such as phosphatidly innositol 3 kinase (PI3K). These labile biological activities are not well suited for correlation to target molecules in our positionally mapped reference database by the assay of biological activity in exclusionary/ inclusionary pairs.

An alternate method of correlation is provided for in the present invention by means of a calibrated plurality of rapid and/or small scale solution based separation means (see FIG. 2) that correspond in the distribution profile to the larger scale solution based separation means used to create the aforementioned reference database of the target set from the particular cell type being assayed. FIG. 2 depicts said plurality of rapid and/or small scale solution based separation means (5,6,7) corresponding to complimentary chromatographic separation columns as described above.

We have depicted three such separation means, however, it is to be understood that the number of said separation means can be greater or fewer than three and ideally should match the number of separation means used to create the corresponding reference database. Said solution based separation means are provided, in the usual manner of chromatographic workstations, with a source of hydraulic flow of a suitable mobile phase (8,9,10) for each separation means and a means of introducing the sample into said mobile phases. In a preferred embodiment of the current invention, said means (11) of introducing the sample into each mobile phase will introduce an equivalent sample into each mobile phase at the same time, thus reducing the time required to separate the sample on the various solution based separation means (5,6,7). A plurality of fractions, collected in arrays(12,13,14), of the elution from each said solution based separation means (5,6,7) are simultaneously and rapidly assayed for the particular biological activity being investigated.

The scale and speed of said separation and assay is such that a profile of the particular biological activity in question can be obtained within the biological activity lifetime of said biological activity. The measured quantity of biological activity in each said fraction in said first array (12) is recorded so as to determine its maxima and distribution profile as depicted in FIG. 3A. The calibrated maxima and distribution profile of each target molecule in the reference database target set corresponding to those fractions that contain biological activity are, in a preferred embodiment, automatically analyzed using a suitable peak analysis computer program such as PeakFit ® (SPSS Inc., Chicago, IL.). Said target molecules represent a subset of the entire target set. The relative fit between the maxima and distribution profile of said biological activity and the maxima and distribution profile of each said target molecule in said target subsets is measured and recorded by any suitable means, such as the

method of sums of squares of the difference between the compared profiles.

For purposes of illustration only, said relative fit values of a plurality of hypothetical target molecules is shown **(15)** vertically plotted as they may have
5   appeared in the aforementioned multi-dimensional analysis method used to analyze and resolve target molecules in the original positionally mapped reference database. Those with the best fit show the highest correlation measurement in the vertical dimension. Likewise in FIGS. 3B and 3C, equivalent biological activity maxima and distribution profiles of the same assayed activity are plotted **(13,14)** and the
10   corresponding relative fit values calculated for target molecules in the various fractions containing biological activity are again plotted **(16,17)**. It should be noted that most of the target molecules in each said correlation are incidentally correlated to the biological activity purely by chance.

15   Calculation of a Boolean intersect of the subset of target molecules in the first subset **(15)** with the subsets of target molecules in the other subsets **(16,17)** is performed (depicted in FIG. 3D for illustration purposes only) wherein the relative fit measurement of equivalent target molecules in the intersecting subsets is summed. The incidentally correlated target molecules in one subset are unlikely to be
20   incidentally correlated in another subset because of the complimentary or orthogonal nature of the various separation means, thus providing for a small group of target molecules (perhaps only one) that are strong candidates for the source or sources of said labile biological activity.

25   The method described provides for the rapid correlation of a measurable labile biological activity to identifiable target molecules in a reference database with only a single and rapid stage of partial purification.

EXAMPLE 3
30

**CORRELATION OF IN VITRO STABLE BIOLOGICAL ACTIVITY
MEASURED IN RECOMBINED INCLUSIONARY/ EXCLUSIONARY POOLS
TO TARGET MOLECULES**

A priori Global Determination of Protein-Small Molecule interactions
Between a cellular Lysate and a Combinatorial Library

Combinatorial libraries containing a multitude of variant small molecules are a
5   valuable new tool for drug discovery. Such combinatorial libraries represent a rich
potential source or compounds, called pharmacophores, that interact with the active
site of enzymes or signaling molecules in living organisms. Global knowledge of the
interaction of any said variant small molecules with particular target molecules in a
cellular lysate is of great potential interest to the drug discovery industry. This
10   information would provide leads for possible pharmacophores within the library before
a particular target is defined. Additionally, potential non-specific interactions as well as
agonist and antagonist relationships between the said small molecules could be
elucidated.

15   A means is provided to globally test a plurality of variant small molecules for
interaction with particular target molecules in a target set such as the entire protein
inventory of a cell lysate using the aforementioned method of formulating pairs of
exclusionary/ inclusionary cell lysate pools. Said method can test said plurality of
variant small molecules one at a time against all said pairs of exclusionary/
20   inclusionary cell lysate pools or more efficiently in groups. Since the rate of positive
interaction with any given pair of said pools is likely to be very low, grouping many
variant small molecules in a single test is desirable. The number of said variant small
molecules tested together should provide a rate of positive interaction of said variant
small molecules and said pairs of exclusionary/ inclusionary cell lysate pools, that is
25   less than one per pair. Preferably, a rate of positive interaction less than one in ten to
one in one hundred pairs is desirable. In order to prevent double positives within a
single test.

A plurality of biologically active or non-denatured target molecules is provided,
30   such as proteins formulated from a plurality of fractions separated by a plurality of
solution base separation means in the manner describe previously in the current
patent. Said formulation provides for said plurality of recombined exclusionary and
inclusionary pools. A test is defined herein as the incubation of one or a plurality of
variant small molecules at one time with two wells of said array containing an
35   exclusionary pool and an inclusionary pool corresponding to one target molecule in

the target set. Many such tests can be performed simultaneously in the current invention.

The method of measuring interaction between said variant small molecules and a said pool will be determined by the nature of the combinatorial library and the manner in which it is formulated. In a preferred embodiment of the current invention the variant small molecules contained in the combinatorial library are labeled with detectable isotopes of the elements found in the combinatorial library. A means of separating interacting small molecules from non interacting small molecules or for measuring proximity between said target molecules and said variant small molecules is provided. Said means include but are not limited to differential filtration, adsorption or sedimentation and proximity scintillation techniques suitable for many small scale tests.

A positive test is determined by the measurement of a relatively strong interaction of the test small molecules with the inclusionary pool and a concomitant measurement of little or no interaction with the exclusionary pool. Measurement of a moderate amount of interaction with both pools is not indicative of a positive result as many target molecules will be incidentally included in a given pair that are not the particular target molecule formulated to be included and excluded in said pair. In the case of a positive test where more than one variant small molecule is tested at one time, additional tests can be performed to identify the particular variant small molecule responsible.

EXAMPLE 4

DETERMINATION OF INTERACTION BETWEEN TARGET MOLECULES WITHIN A TARGET SET RESULTING IN ENHANCED BIOLOGICAL ACTIVITY

Discovery of Kinase-Substrate relationships

In the aforementioned prior art of two and three hybrid measurements of protein-protein interaction in which the test pair are expressed in a recombinant vector and interaction is detected by some measure of proximity, the biological significance

of the so measured interaction is only suggestive. In most such surrogate conditions, the recombinantly expressed target protein is not post-translationally modified or processed in the same manner that it is in the native state. A test of the target protein in its native state provides a global means of measuring actual biological interaction

5   between target molecules in a target set such as the complete cellular protein inventory of a cell type. Said test of the target protein in its native state is provided for with the aforementioned method of pairs of exclusionary/ inclusionary cell lysate pools

An example of a well-known general biological interaction between proteins in

10   a cellular inventory is the covalent phosphorylation of one protein by another. An enzyme that phosphorylates another substrate protein is called a kinase. Such interaction is often a signal transfer step involved in a signal cascade in which a small initial signal such a receptor activation is propagated and amplified. In order to provide signal fidelity, the phosphorylation enzyme-substrate relationships between proteins

15   are known to be very specific and depend on interaction between particular pairs of kinases and substrates. Discovering the specific kinase-substrate relationships within a target set such as the complete cellular inventory of proteins is a goal of the current example.

20   In this example, a means is provided to globally test in vitro a plurality of target molecules within a target set, such as the entire protein inventory of a cell lysate, for kinase-substrate relationships under native conditions. An extension of the method of formulating pairs of exclusionary/ inclusionary cell lysate pools provides a means of testing all target molecules in a target set against all others for enzyme-

25   substrate relationships, in this case the ability to covalently attach a radio-labeled phosphate group to some target molecule within a test pair of said recombined cell lysate pools.

In the current example a means is provided to formulate an array of doubly

30   inclusive and exclusive cell lysate pools using robotic liquid handling apparatus in which (in contrast to the previously described means) not one but two target molecules from a target set are included and excluded by careful pooling of fractions from the same said solution based separation means used to create single target molecule inclusive and exclusive pools. As described for the formulation of said

35   single target molecule inclusive and exclusive pools, said doubly inclusive and

exclusive cell lysate pools are created utilizing the target positional information provided by the aforementioned comprehensive multi-dimensional mapping reference database. A plurality of doubly inclusive and exclusive cell lysate pools are subsequently tested in the current example by incubation of said doubly inclusive and

5   exclusive cell lysate pools with a suitable phosphate substrate for said kinases that can be incorporated into a substrate target molecule in the target set.

A positive test is defined, in the current example, as the determination of covalently bound phosphate in said doubly inclusive cell lysate pool with a

10   concomitant reduction or total lack of covalently bound phosphate in said doubly exclusive cell lysate pool. As before, a moderate signal in both said pools is inconclusive and does not constitute a positive test. In a preferred embodiment of the current invention, said phosphate substrate is provided as radiolabeled phosphorous in the γ phosphate position of adenosine triphosphate that is detectable by existing

15   means of radioactive detection. A means is provided to segregate unincorporated adenosine triphosphate from covalently bound phosphate. Said means include but are not limited to differential filtration, adsorption or sedimentation and proximity scintillation techniques suitable for many small scale tests.

20   A related method to quickly optimize the discovery of said kinase-substrate relationships in complex target sets such as the complete cellular protein inventory of a cell type is provided. Said related method narrows candidates for said kinase-substrate relationships to subsets of the overall target set. This allows for subsequent application of the aforementioned method of doubly exclusive and inclusive pools to

25   elucidate the particular target molecule within said target molecule subset only.

In target sets such as the complete cellular protein inventory of a cell type the number of pairs of target molecules combined in doubly exclusive and inclusive pairs can be impractical to test pair-wise two at a time. Also, the number of said positive

30   tests can be inefficiently low. Said related method of formulating recombined cell lysate pools provides for an initial low stringency test in which a plurality of target molecules are identified as a subset of the complete target set. Again, utilizing the information in the aforementioned reference database, robotic liquid handling apparatus recombines said fractions from solution based separation means so as to

35   formulate a pool of fractions that specifically excludes all said plurality of target

molecules of said subset. Likewise a recombined pool of fractions from solution based separation means is formulated that specifically includes all said plurality of target molecules. A means is provided, as before, to test the exclusionary and inclusionary pools corresponding to said subset of said plurality of target molecules.

5

A positive result, in the current example, is defined as the determination of covalently bound phosphate in said subset inclusive cell lysate pool with a concomitant reduction or total lack of covalently bound phosphate in said subset exclusive cell lysate pool. As before, a moderate signal in both said pools is

10 inconclusive and does not constitute a positive test. A positive test for a given subset of a plurality of target molecules provides information that one or more pairs of target molecules within said subset plurality of target molecules have a kinase-substrate relationship without determining which pairs are responsible. Said positive test allows for formulating doubly exclusive and inclusive pools with those target molecules within

15 the positive testing subset only. Said formulation of said subset exclusive and inclusive pools allows for additional testing with the method of said doubly exclusive and inclusive pools to eliminate ambiguity about the identity of the pair or pairs of target molecules within said subset responsible for said positive test.

20 EXAMPLE 5

DETERMINATION OF GLOBAL INTERACTION BETWEEN A STOCHASTICALLY GENERATED LIBRARY AND A TARGET SET RESULTING IN A PLURALITY OF BIOLOGICAL ACTIVITIES

25

Discovery of a Plurality of Candidate Molecules for Binding Epitopes or Substrates to target Molecules in a Target Set.

A stochastically generated plurality of expressed peptide sequences that

30 provide a means of correlation of the said peptide sequence to the genetic sequence encoded in the vector that expresses said peptide sequence is provided. Each said sequence is by some means displayed on the surface of said genetic vector responsible for said peptide sequence. Said vectors are generally known as surface display expression vectors. The screening of said surface display expression vectors

to identify ligands for target molecules such as proteins followed by a process of clonal expansion of said vectors with variation is generally known as directed molecular evolution.

5   A means is provided in the present invention to screen and discover a plurality of stochastically generated surface display expression vectors, that interact through the specific variant peptide sequences so displayed, with specific target molecules in a target set by any means of biological assay that is stable in vitro in a recombined pool of cellular fractions, wherein the said specific variant peptide sequence causes a

10   measurable biological effect. In the present example, a means for selecting and screening stochastically generated candidate molecules from stochastic libraries of said surface display expression vectors that interact by binding to target molecules within aforementioned exclusionary and inclusionary pools is provided. In addition, in the present example, a means for selecting and screening stochastically generated

15   candidate molecules from stochastic libraries of said surface display expression vectors that interact so as to chemically modify said stochastically generated candidate molecules in a detectable manner is provided.

In the current example in which ligand interaction is the particular biological

20   activity tested for, a positive test is determined by the preincubation of said plurality of stochastically generated surface display expression vectors with the exclusionary pool of a pair of exclusionary-inclusionary pools of target molecules. The subset of preincubated stochastically generated surface display expression vectors that do not interact and are unbound to any target molecules within the exclusionary pool is

25   recovered by any of the aforementioned means of bound to unbound separation. Said recovered subset of preincubated stochastically generated surface display expression vectors is subsequently incubated with said inclusionary pool of target molecules. A means of separating bound from unbound stochastically generated surface display expression vectors in said inclusionary pool target set is provided.

30

Said positive test, in the current example, is represented by the binding of one or a plurality of surface display expression vectors to target proteins in the inclusionary pool to form a second subset of surface display expression vectors that are candidates for specific interaction with the particular target molecule for which the

35   exclusionary and inclusionary pools were formulated. It should be noted that said

second subset represents an enriched subset of potential ligands for clonal expansion and further rounds of screening and discovery. Those potential ligands so discovered are candidate molecules for specific interaction with said particular target molecule for which the exclusionary and inclusionary pools were formulated in the presence of all

5   the other target molecules within the target set. Thus a means is provided to simultaneously screen and discover potentially interactive molecules for every pair of exclusionary and inclusionary pools in a given target set through a plurality of rounds of clonal expansion of said subsets of surface display expression vectors such that non-specific or cross-reactive interactions are selected against.

10

If said test interaction between surface display expression vectors and exclusionary and inclusionary pools formulated for a particular target molecule involves a chemical modification of the variant peptide of said surface display expression vectors, a means is provided to recover and detect said chemical

15   modification after interaction with a test exclusionary and inclusionary pool. For example, screening and discovery of variant peptides of surface display expression vectors that function as potential specific exogenous kinase substrates for particular target molecules is provided. Incorporation of radiolabeled phosphate into the surface display expression vector is measured and detected as described above for

20   endogenous target-target interaction.

A positive test, indicating a potentially specific kinase substrate for a particular target molecule involves, first, a measurement of radiolabeled phosphate incorporation into said surface display expression vectors following incubation with the

25   inclusionary pool formulated for said particular target molecule. Second, it involves a measurement of little or no radiolabeled phosphate incorporation into said surface display expression vectors following incubation with the exclusionary pool formulated for said particular target molecule. As before, non-differential measurements in both exclusionary and inclusionary pools are inconclusive and represent a negative test

30   result. Additional tests following clonal expansion with variation are greatly enhanced by the selection and separation of said radiolabeled surface display expression vectors. In a preferred embodiment, said phosphorylated surface display expression vectors are separated from non-phosphorylated surface display expression vectors using the known art of immunoaffinity separation with phosphorylated peptide specific

35   antibodies.

The current examples of screening and discovery of interactive peptides form stochastically generated surface display expression vectors that interact with a complete target set provide for the identification of initial candidate molecules only. A
5   program of simultaneous directed molecular evolution or directed molecular co-evolution requires a further means of scoring said initial candidate molecules for subtle differences in their interaction with a particular target molecule and for subtle differences in their unwanted interaction with non-target molecules.

10   EXAMPLE 6

SCORING OF A PLURALITY OF INITIALLY SELECTED CANDIDATE MOLECULES FOR HIGHLY SPECIFIC INTERACTION WITH MOLECULES IN A TARGET SET.

15

A means is provided to partition a plurality of related or unrelated initially selected candidate molecules in a stochastically generated library of surface display expression vectors on the basis of subtle differences of interaction with said particular target molecule for which they were initially selected and for subtle differences of
20   interaction with said non-target molecules for which they were initially selected against.

A means is provided to covalently immobilize said variant peptide sequences, as displayed on said surface display expression vectors, or as isolated peptides onto
25   a paramagnetic particle having dimensions in the nanometer to micrometer range. Said paramagnetic particle will move by magnetic force in an externally applied magnetic force field. A plurality of initial candidate molecules are so immobilized in separate immobilization reactions such that a single candidate molecule is present on the paramagnetic particles within a compartment. After said separate immobilization
30   reactions, said paramagnetic particles with all initial candidate molecules are mixed to form a single pool of paramagnetic particles. Said pool of paramagnetic particles can be subdivided to provide multiple test pools to the extent that each initial candidate molecule is represented in each said subdivided multiple test pool.

Said single pool of paramagnetic particles is incubated with an exclusionary target set pool formulated as described above to contain all target molecules except the particular target molecule for which the initial candidate interactive molecule has been selected. Possible non-specific or low affinity interaction of said immobilized

5   candidate molecule with non-target molecules in said exclusionary target set pool results in a loosely associated molecular complex surrounding the paramagnetic particle. Said paramagnetic particle and its immobilized candidate ligand is partitioned between a stationary phase, such as a porous matrix, by differences in the resistance to an applied magnetic force field cause by a differences in the partition coefficient of

10   said paramagnetic particles. In a preferred embodiment of the current invention said partition is provided by differences in the mobility of said paramagnetic particle due to size or steric hindrance of any said loosely associated molecular complex. Paramagnetic particles having little or no interaction with non-target molecules in said target set will exhibit the least resistance to movement by the said externally applied

15   magnetic force field and will thus segregate ahead of the paramagnetic particles that do exhibit non-specific interaction, however subtle, with the non-target molecules in said target set.

A further means of segregating or fractionating the paramagnetic particle

20   stream so as to isolate the various paramagnetic particles with immobilized initial candidate molecules is provided. Information is thus determined to allow the relative scoring of specificity towards a particular target molecule of an initial variant candidate molecule from a stochastically generated library. Variant candidate molecules in additional rounds of clonal expansion with variation can also be scored in this fashion.

25

EXAMPLE 6

A MEANS OF DISCOVERY AND SCREENING OF A PLURALITY OF INITIAL CANDIDATE MOLECULES FOR BINDING EPITOPES TO
30   ## ISOLATED IMMOBILIZED TARGET MOLECULES IN A TARGET SET

The previous examples of discovery of candidate molecules from a stochastically generated library of surface display expression vectors generally involve formulation of target molecule pools in solution. By contrast, an efficient

means is provided to select initial candidate binding ligands to all target molecules in a target set wherein the particular target molecule is isolated and immobilized onto a paramagnetic particle. Said paramagnetic particle is preferably in the nanometer size range, but in a size range distinguishable from the size of the said surface display

5   expression vector used to generate said stochastic library. A source of isolated target molecules or fragments of said isolated target molecules such as proteins and peptide fragments of isolated target molecules is provided. Said isolated target molecules are separately immobilized onto paramagnetic particles using any means of covalent attachment.

10

A stochastically generated library of candidate ligands expressed on the surface of a micrometer scale surface display expression vector such as a recombinant yeast vector is provided. Preferably, said stochastically generated surface display expression vectors are divided during construction into a plurality of

15   sub-libraries containing unique stochastic subsets. As depicted schematically in FIG. 4A, each said sub-library (18) is isolated into a single compartment or well in a micro-array plate (19).

A means of rapid and efficient massively parallel discovery of initial candidate

20   ligands to all target molecules in a target set is provided. Said isolated target molecules (20) are separately immobilized onto paramagnetic particles (21). Said paramagnetic particles are incubated in a single said sub-library (18) of said stochastically generated surface display expression vectors (22). As shown in FIG. 4B, following equilibration of any potential ligand interaction for a pre-determined time,

25   a microporous screen (23) is placed over said micro-array plate (19). Said microporous screen hole size is such that the said paramagnetic particles (21) can easily pass through but said surface display expression vector (22) is completely retained. A second micro-array plate (24) having additional sub-libraries is placed over said incubated micro-array plate (19) containing the equilibrated incubation test

30   such that the wells in one micro-array plate align with the wells in the other micro-array plate. A magnetic force field (25) is externally applied perpendicular to the faces of the combined micro-array plates as shown in FIG. 4C so as to move the non-interacting paramagnetic particles (21) into the second micro-array plate (24). Interacting paramagnetic particles (26) bound to surface display expression vectors

35   (27) will be held back from transfer and retained in said first micro-array plate (19)

because the interacting surface display expression vector (27) cannot pass through said microporous screen (23). In FIG. 4D said magnetic force field (25) is removed and said first micro-array plate (19) is removed from the assembly (28) of said microporous screen (23) and said second micro-array plate (24). A third empty micro-

5   array plate (29) as shown in FIG.4E, for collecting any interacting magnetic particles is placed over said first micro-array plate (19) without any microporous screen, again with alignment of wells between the two plates. A magnetic force field (25) is again externally applied perpendicular to the faces of the combined micro-array plates (19, 29) so as to move the interacting paramagnetic particles (26) and any attached

10   surface display expression vectors (27) into said empty third micro-array plate (29). As shown in FIG. 4F, said magnetic force field (25) is removed and said third micro-array plate (29) containing potential candidate surface display expression vector (27) is analyzed for the presence of a bound paramagnetic particle (26). In a preferred embodiment of the invention the paramagnetic particle (26) is detected by a

15   magnetometer (30) in a weak magnetic field such as the Earth's magnetic field. An example of a suitable magnetometer is a super-conducting quantum interference device.

The non-interacting paramagnetic particles (21) of the first test are

20   subsequently incubated and tested in a like manner in the sub-library of the second micro-array plate (24). In this manner a plurality of target molecules are serially passed from one sub-library of surface display expression vectors to another to test for interaction. A plurality of target molecules bound to paramagnetic particles are thus tested in a massively parallel fashion for discovery of interaction with some initial

25   stochastically generated candidate molecule in a specific sub-library. Additionally, the initial candidate molecule is segregated from said sub-library and collected for subsequent rounds of clonal expansion and interaction with said particular target molecule.

30   A means is provided to allow every target molecule to incubate and be tested against every said sub-library in a serial manner. In order to increase the number of positive tests to approximately one in ten tests or one in a hundred tests, combinatorial sets of unrelated target molecules on said paramagnetic particles are mixed in a plurality of semi-replicate pools such that no two replicate pools contain

35   more than one common target molecule. A positive test will reveal a pattern of

replicate positives in a subset of said replicate pools corresponding to the subset of pools containing the source of the positive test. In this manner the identity of the particular positive target molecule is revealed by the combinatorially determined intersects of the various replicate pools.

5

Using this method, the number of combined tests can be greatly increased. It should be noted, however, that more than one positive result within a single combined test will be ambiguous as to the identity of the interacting target due to the detection of more than the standard number of positives within the replicate pool set. Said
10   ambiguity is minor and easily eliminated by further testing.

15

We claim:

1. A method for separating, characterizing, screening, and identifying biological
    targets comprising the steps of:

    (a) integrating biological material into said gels;

    (b) characterizing said biological material comprising running said integrated
        gels through two-dimensional electrophoresis in order to obtain a first and
        second identification parameter;

    (c) further characterizing said biological material in order to obtain a third
        identification parameter;

    (d) plotting said first, second, and third identification parameters to generate a
        library of biological targets having identification parameters for said
        biological material;

    (e) pre-fractionating by a sub-cellular isolation means or physiochemical
        criteria allowing a sufficient amount of mass to be loaded on each solution
        based separation means to detect lowest abundance biological molecules
        in a target set;

    (f) determining a set of positional coordinates for biological molecules where
        given molecules have unique positional coordinates;

    (g) using said positional coordinates as information to recombine a plurality of
        isolated fractions to form carefully formulated pools created for each target
        molecule in a target set;

    (h) correlating measurable labile biological activity to identifiable target
        molecules in a reference database with only a single and rapid stage of
        partial purification;

    (i) pair-wise screening of target molecules for biologically significant target-
        target interactions by combinatorial recombination of fractions;

    (j) pre-screening for interaction, to a predetermined level, against an
        exclusionary pool formulated to exclude a particular target molecule;

    (k) pre-screening for interaction against an inclusionary pool formulated to
        enhance target molecules;

    (l) screening candidate molecules from a library;

    (m) screening said library with ligands;

    (n) detecting said ligand-biological target interactions;

    (o) scoring the relative interactions of candidate molecules against a chosen

target and scoring the relative interaction of candidate molecules for
unwanted interaction with all other target molecules in a target set;

(p) partitioning related candidate molecules attached directly or indirectly to
magnetic particles; and

(q) identifying initial candidate molecules from a library of surface display
expression vectors wherein said target molecules or fragments of said
target molecules in a purified form are attached to magnetic particles
having a smaller size distinguishable from a size of a group of vectors in
said library.

2. The method according to claim 1, step (g) further comprising the step of
recombining said isolated fractions to create a subset pool of an entire target set
containing a given target in an enriched manner.

3. The method according to claim 1, step (g) further comprising the step of
recombining said isolated fractions that do not contain any trace of a given target
molecule, creating a subset pool of an entire target set that incidentally includes
about every molecule in said target set except said target molecule.

4. The method according to claim 1, step (o) further comprising the step of scoring
relative interactions among closely related variant candidate molecules by
attaching said related candidate molecules or the vectors that display said related
candidate molecules to a plurality of magnetic particles, having a size of about
one nanometer to about one micrometer, and allowing said magnetic particles to
interact in free solution with an exclusionary pool or an inclusionary pool so as to
bind those target molecules that interact with said related candidate molecules.

5. The method according to claim 1, step (p) further comprising the step of placing
said magnetic particles in a mobile solution phase and drawing it through a
stationary phase medium by application of a magnetic field.

6. The method according to claim 1, step (p) further comprising the step of using
competitive interaction of target molecules in solution with magnetic particles
having related candidate molecules attached and a stationary matrix having
relatively weak interacting candidate molecules or vectors immobilized on a matrix

surface.

7. An apparatus for separating, characterizing, screening, and identifying biological
    targets comprising:

    (a) a means for continuously producing uniformly-formed, polymerized, and
        cut gels;

    (b) a means for integrating biological material into said gels;

    (c) a means for characterizing said biological material comprising running said
        integrated gels through two-dimensional electrophoresis in order to obtain
        a first and second identification parameter;

    (d) a means for further characterizing said biological material in order to obtain
        a third identification parameter;

    (e) a means for plotting said first, second, and third identification parameters
        to generate a library of biological targets having identification parameters
        for said biological material, which allows for the identification of said
        biological targets;

    (f) a means for positionally mapping every protein in an aggregate, such as a
        cell lysate, allowing for the selective recombining by pooling of semi-
        purified cellular fractions profiled on multiple separation means;

    (g) a means for screening said library with ligands; and

    (h) a means for detecting said ligand-biological target interactions.

8. The apparatus according to claim 7, further comprising, a means for screening,
    selecting, scoring and characterizing a level of biological activity interactions with
   many target molecules in a cell in parallel.
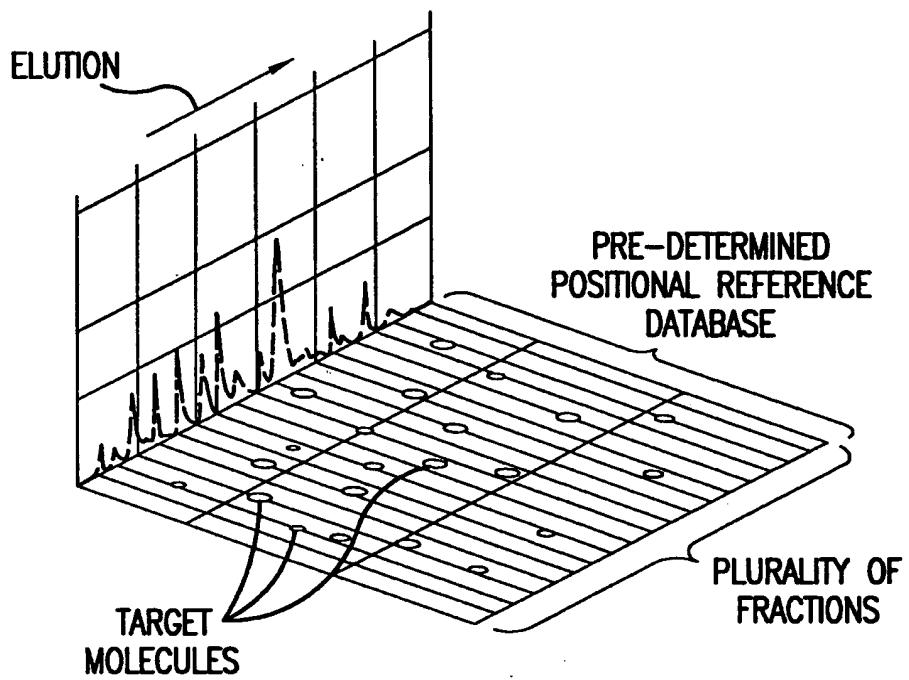
FIG.1A

FIG.1B

ELUTION

PROFILE A
PROFILE B
PROFILE C
PROFILE D
PROFILE E
PROFILE F

4

4

3

FIG.1C

FIG. 2

FIG.3A

FIG.3B
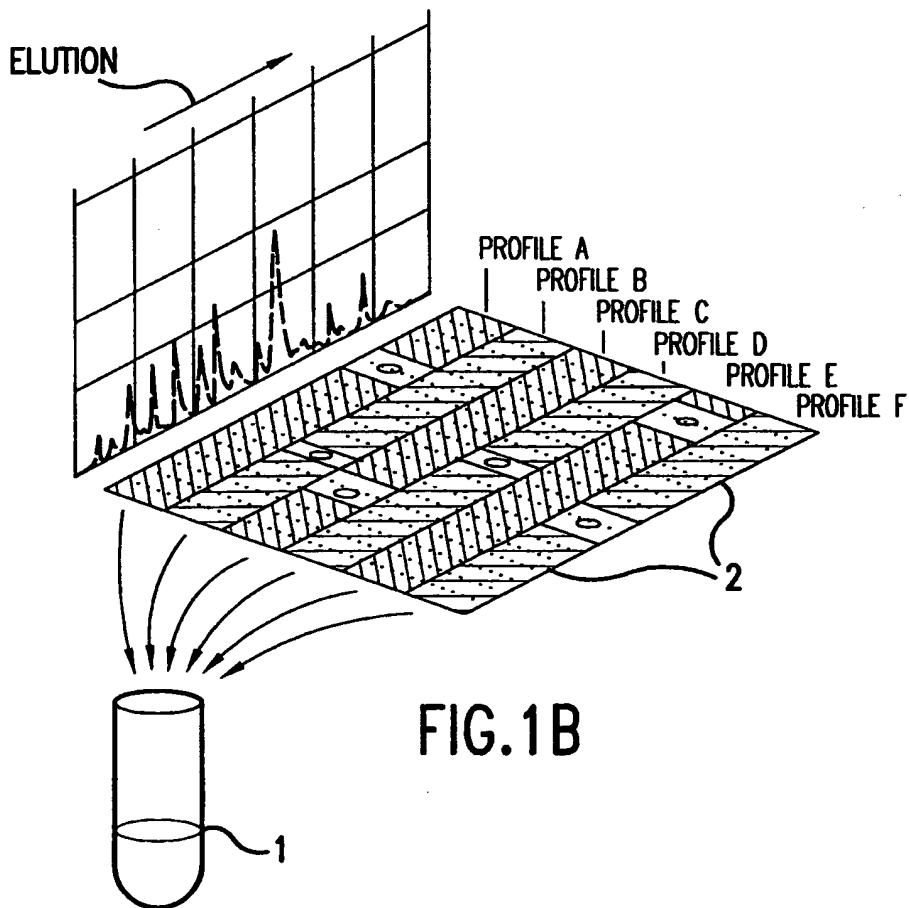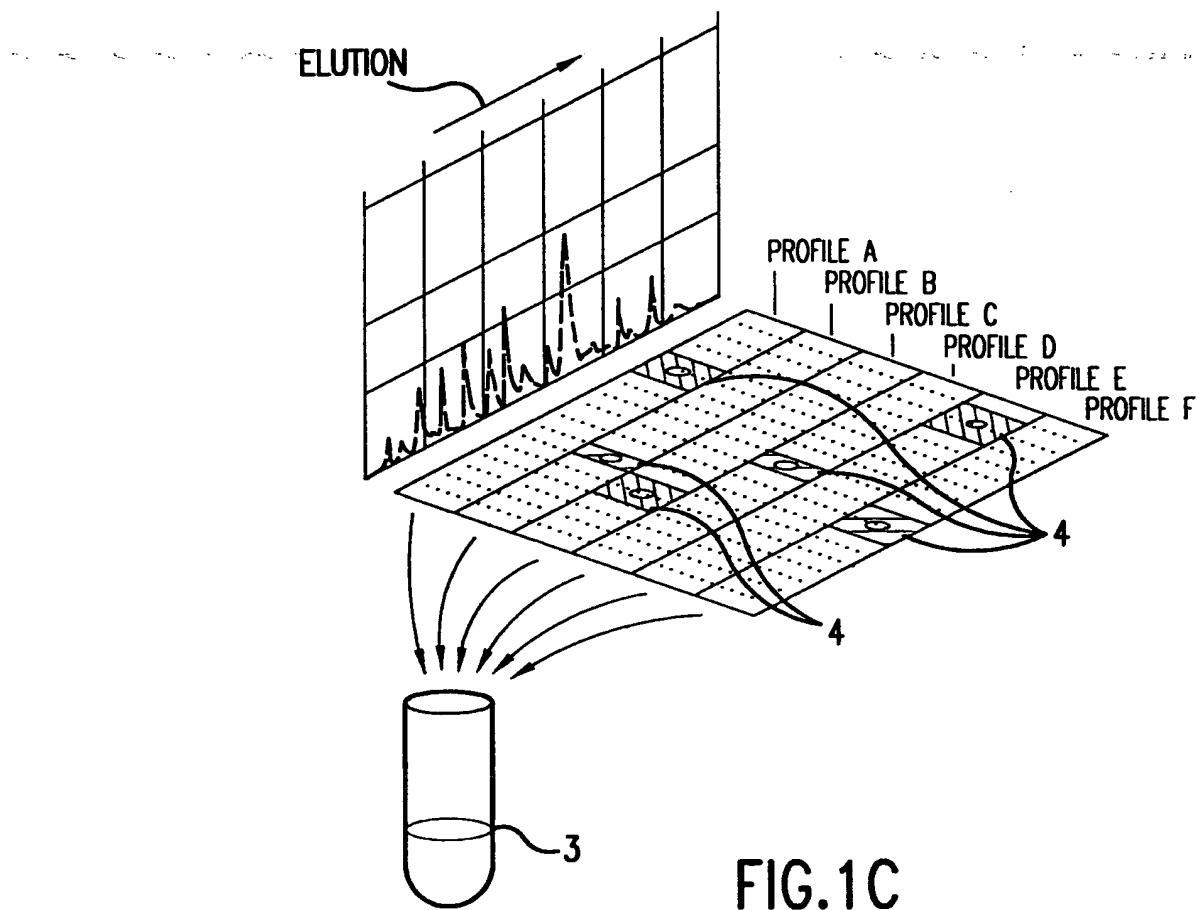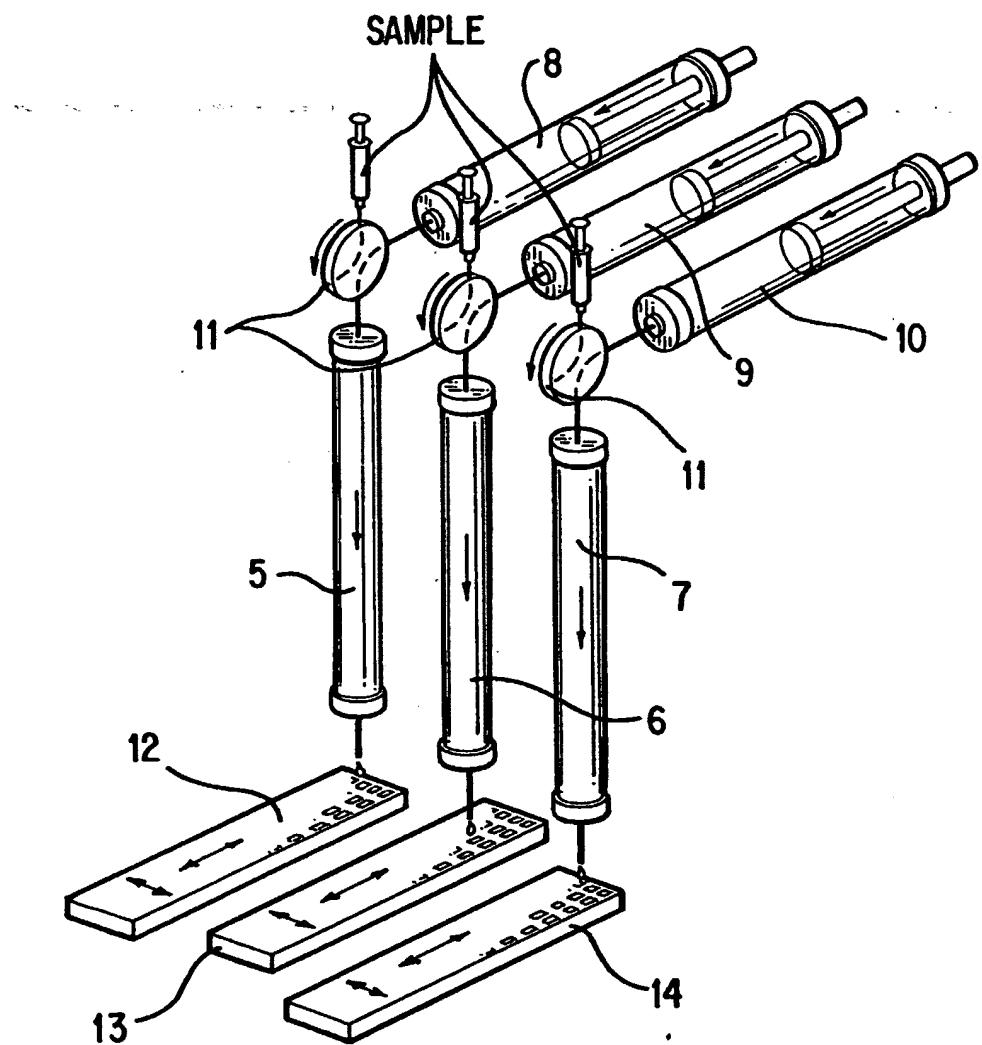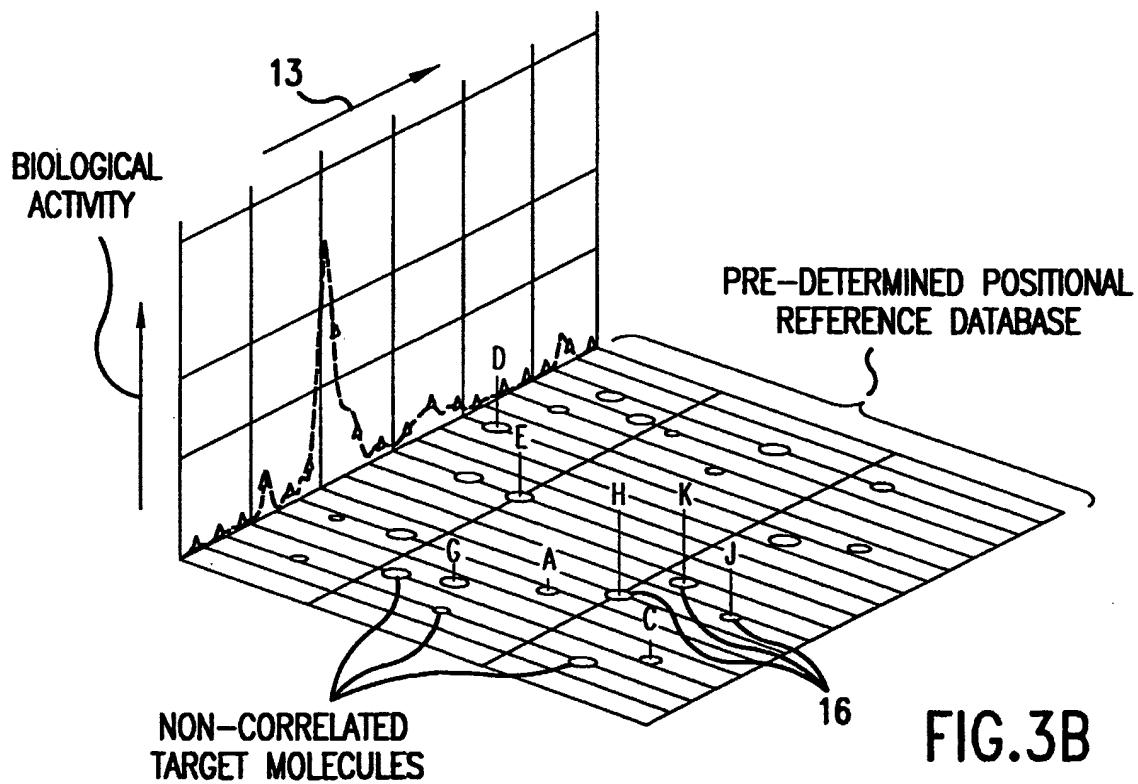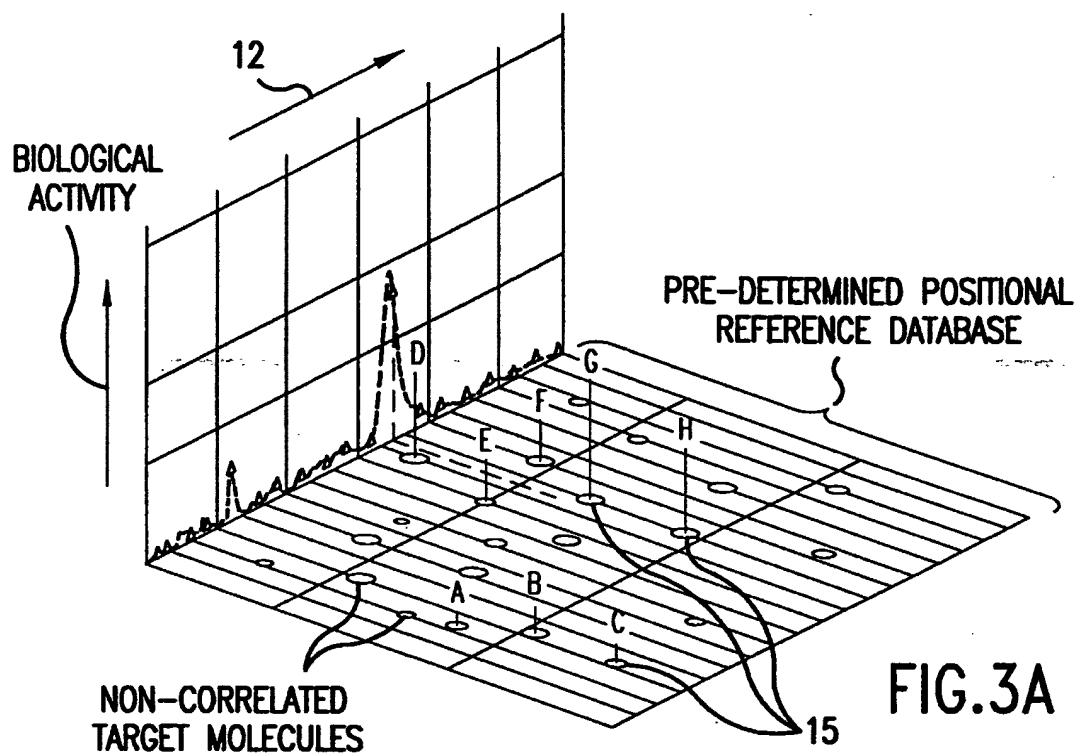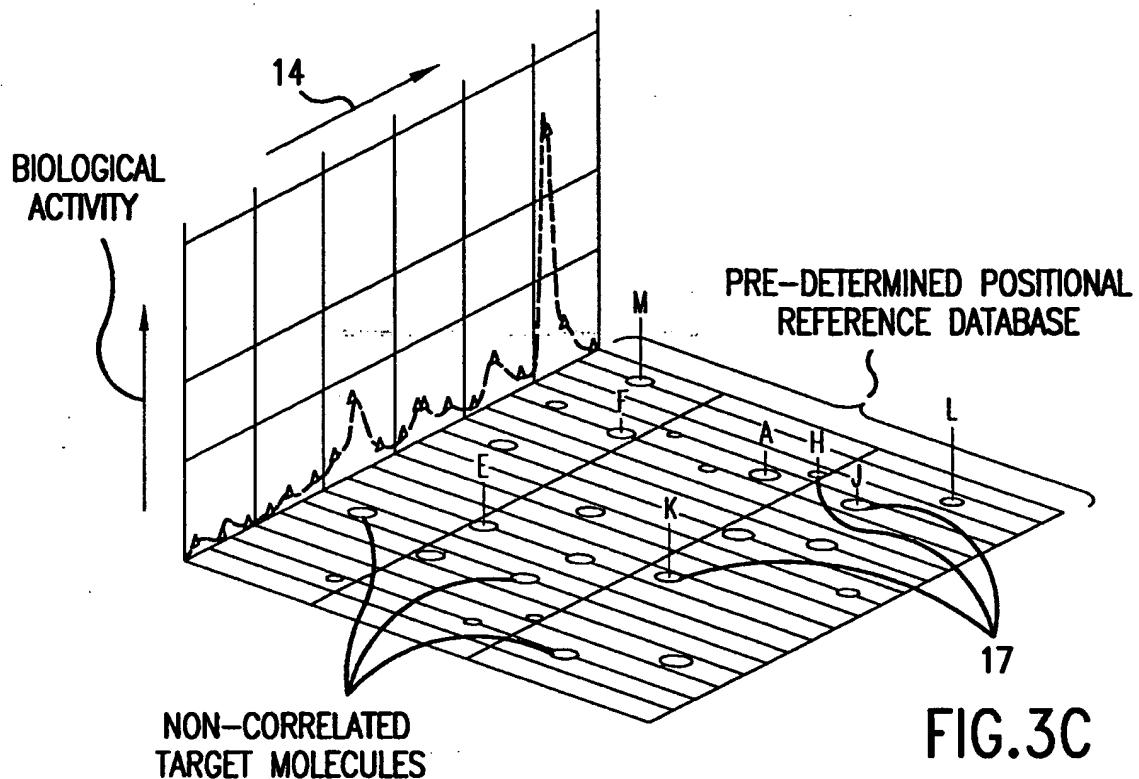
FIG.3C
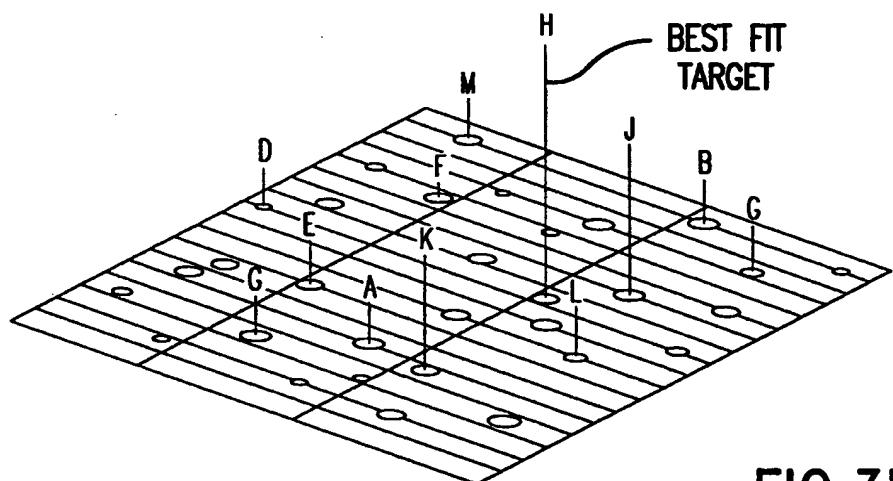
FIG.3D
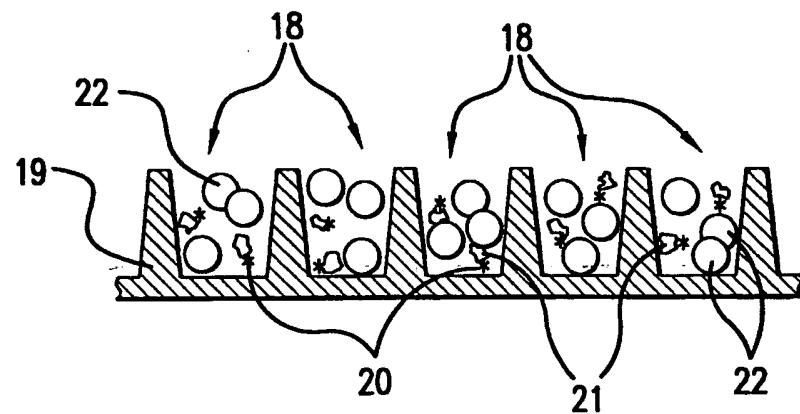
FIG.4A
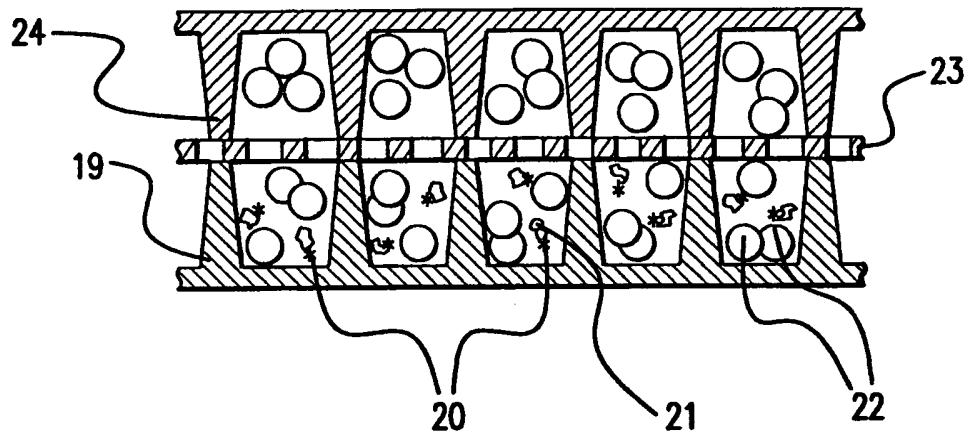
FIG.4B
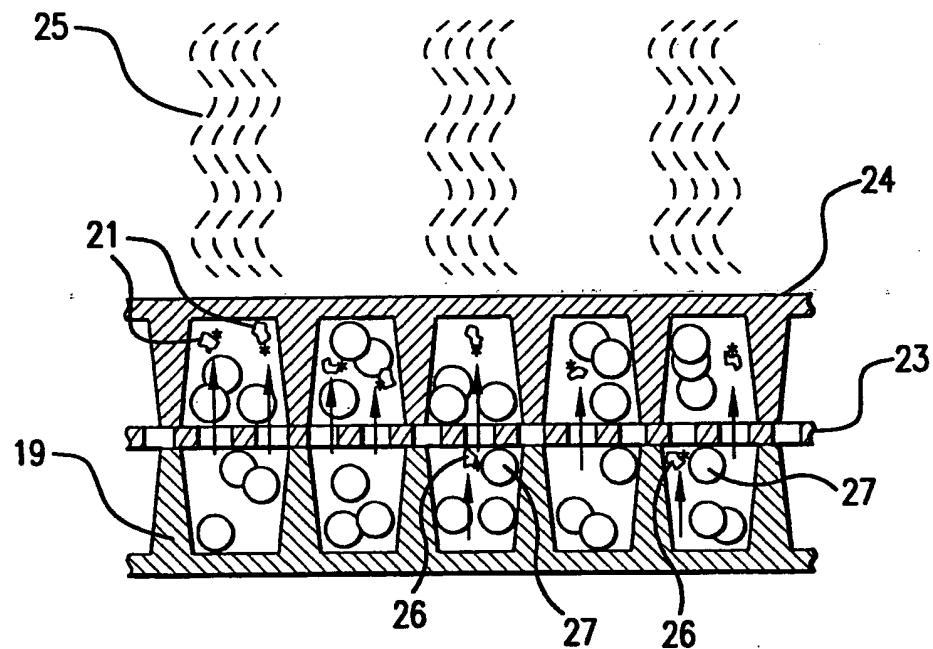
FIG.4C
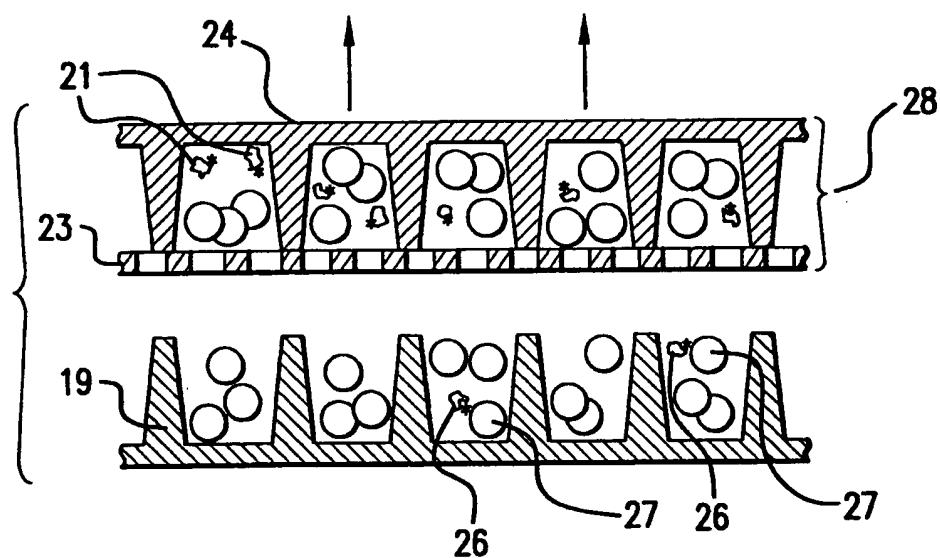
FIG.4D
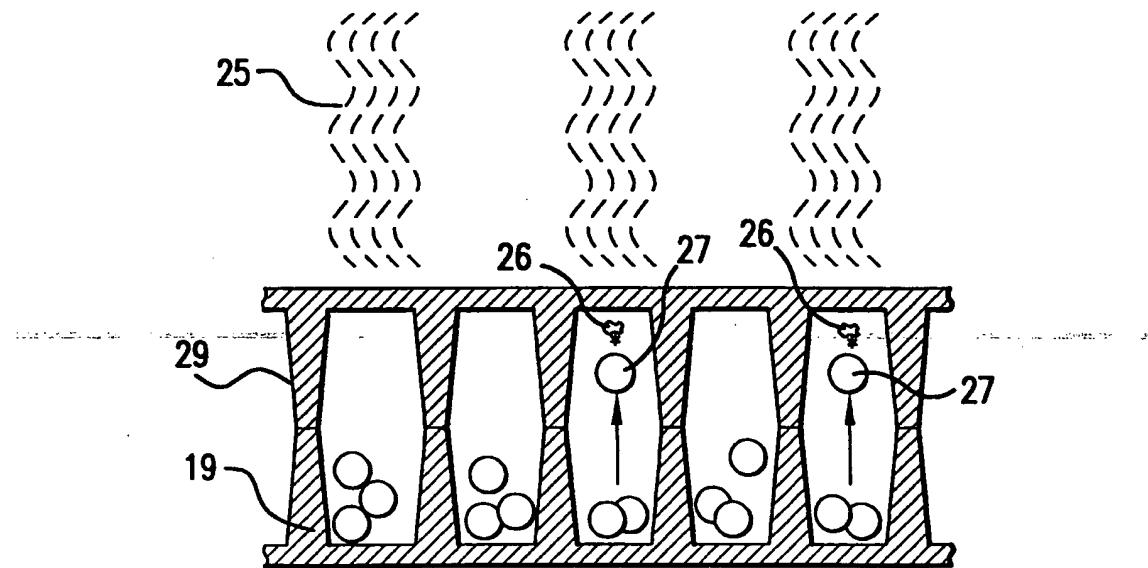
FIG.4E
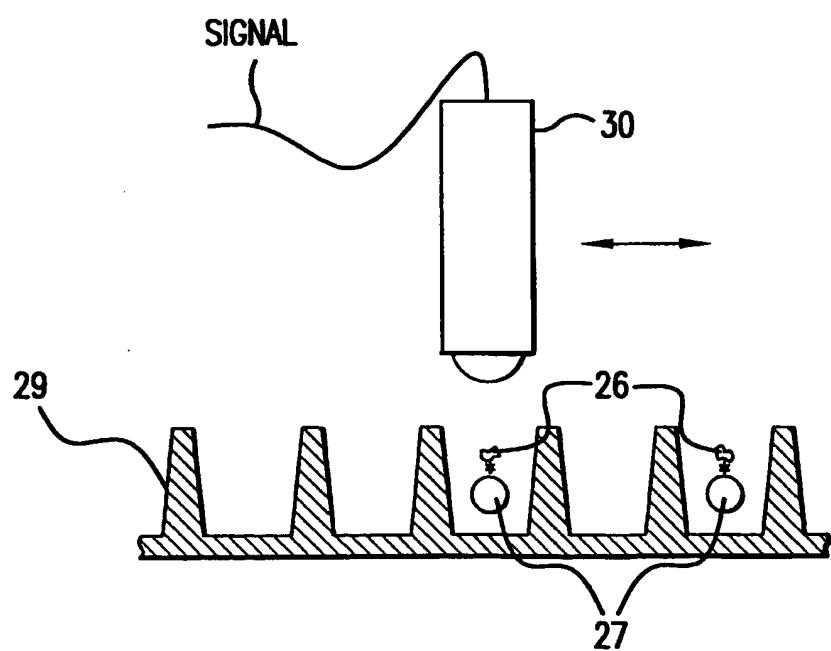
FIG.4F

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :G01N 33/543; C12Q 1/68
US CL : 436/518; 435/6; 204/450, 456, 465, 606

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 436/518; 435/6; 204/450, 456, 465, 606

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

MEDLINE, SCISEARCH, BIOSIS, EAST, WEST 1.2

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WILKINS et al. From Proteins to Proteomes: Large Scale Protein Identification by Two-Dimensional Electrophoresis and Amino Acid Analysis. Bio/Technology. January 1996, vol. 14, pages 61-65, entire document.	1-6
Y	US 4,790,919 A (BAYLOR, JR.) 13 December 1988, see entire document.	7-8
Y, P	GYGI et al. Correlation Between Protein and mRNA Abundance in Yeast. Molecular And Cellular Biology. March 1999, Vol. 19, No. 3, pages 1720-1730, entire document.	1-6

[X] Further documents are listed in the continuation of Box C.   [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 FEBRUARY 2000	17 MAR 2000

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	P. PONNALURI
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196
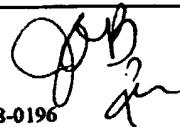
Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y, P	BINZ et al. A Molecular Scanner To Automate Proteomic Research and To Display Proteome Images. Anal. Chem. 01 November 1999, Vol. 71, No. 21, pages 4981-4988, entire document.	1-6
Y.	FELICI et al. Phage-Displayed peptides as Tools for Characterization of Human Sera. Methods in Enzymology. 1996, Vol. 267, pages 116-129, entire document.	1-6